UNIT 1 INTRODUCTION TO STATISTICS

Structure

1.0 Introduction

1.1 Objectives

1.2 Meaning of Statistics

- 1.2.1 Statistics in Singular Sense
- 1.2.2 Statistics in Plural Sense
- 1.2.3 Definition of Statistics

1.3 Types of Statistics

- 1.3.1 On the Basis of Function
- 1.3.2 On the Basis of Distribution of Data

1.4 Scope and Use of Statistics

1.5 Limitations of Statistics

1.6 Distrust and Misuse of Statistics

1.7 Let Us Sum Up

1.8 Unit End Questions

1.9 Glossary

1.10 Suggested Readings

1.0 INTRODUCTION

The word statistics has different meaning to different persons. Knowledge of statistics is applicable in day to day life in different ways. In daily life it means general calculation of items, in railway statistics means the number of trains operating, number of passenger's freight etc. and so on. Thus statistics is used by people to take decision about the problems on the basis of different type of *quantitative and qualitative* information available to them.

However, in behavioural sciences, the word 'statistics' means something different from the common concern of it. Prime function of statistic is to draw statistical inference about population on the basis of available quantitative information. Overall, statistical methods deal with reduction of data to convenient descriptive terms and drawing some inferences from them. This unit focuses on the above aspects of statistics.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Define the term statistics;
- Explain the status of statistics;
- Describe the nature of statistics;
- State basic concepts used in statistics; and
- Analyse the uses and misuses of statistics.

SHOUT
SHARE
IT

1.2 MEANING OF STATISTICS

The word statistics has been derived from Latin word 'status' or Italian 'Statista' meaning statesman. Professor Gott Fried Achenwall used it in the 18th century. During early period, these words were used for political state of the region. The word 'Statista' was used to keep the records of census or data related to wealth of a state. Gradually, its meaning and usage extended and thereonwards its nature also changed.

The word statistics is used to convey different meanings in singular and plural sense. Therefore it can be defined in two different ways.

1.2.1 Statistics in Singular Sense

In singular sense, 'Statistics' refers to what is called statistical methods. It deals with the collection of data, their classification, analysis and interpretations of statistical data. Therefore, it is described as a branch of science which deals with classification, tabulation and analysis of numerical facts and make decision as well. Every statistical inquiry should pass through these stages.

1.2.2 Statistics in Plural Sense

'Statistics' used in plural sense means that quantitative information is available called 'data'. For example, information on population or demographic features, enrolment of students in Psychology programmes of IGNOU, and the like. According to Websters "Statistics are the classified facts representing the conditions of the people in a State specifically those facts which can be stated in number or in tables of number or classified arrangement".

Horace Secrist describes statistics in plural sense as follows : " By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standard of accuracy , collected in a systematic manner for a pre-determined purpose and placed in relation to each other." Thus Secrist's definition highlights following features of statistics:

i) *Statistics are aggregate of facts:* Single or unrelated items are not considered as statistics.

ii) *Statistics are affected by multiplicity of causes:* In statistics the collected information are greatly influenced by a number of factors and forces working together.

iii) *Statistics are numerical facts:* Only numerical data constitute statistics.

iv) *Statistics are enumerated or estimated with a reasonable standard of accuracy:* While enumerating or estimating data, a reasonable degree of accuracy must be achieved.

v) *Statistics are collected in a systematic manner:* Data should be collected by proper planning by utilising tool/s developed by trained personnel.

vi) *Statistics are collected for a predetermined purpose :* It is necessary to define the objective of enquiry, before collecting the statistics. The objective of enquiry must be specific and well defined.

vii) *Statistics should be comparable:* Only comparable data will have some meaning. For statistical analysis, the data should be comparable with respect to time, place group, etc.

Thus, it may be stated that " All statistics are numerical statements of facts but all numerical statements of facts are not necessarily statistics ".

1.2.3 Definition of Statistics

In this unit emphasis is on the term statistics as a branch of science. It deals with classification, tabulation and analysis of numerical facts. Different statistician defined this aspect of statistics in different ways. For example.

A. L. Bowley gave several definitions of Statistics:

i) "Statistics may be called the science of counting" . This definition emphasises enumeration aspect only.

ii) In another definition he describes it as " Statistics may rightly be called the science of average".

iii) At another place Statistics is defined as, "Statistics is the science of measurement of social organism regarded as a whole in all its manifestations".

All three definitions given by Bowely seem to be inadequate because these do not include all aspects of statistics.

According to **Selligman** "Statistics is the science which deals with the methods of collecting, classifying, presenting , comparing and interpreting numerical data collected to throw some light on any sphere of enquiry".

Croxton and Cowden defined "statistics as the collection , presentation, analysis ,and interpretation of numerical data".

Among all the definitions , the one given by Croxton and Cowden is considered to be most appropriate as it covers all aspects and field of statistics.

These aspects are given below:

Collection of Data : Once the nature of study is decided , it becomes essential to collect information in form of data about the issues of the study. Therefore, the collection of data is the first basic step. Data may be collected either from primary source or secondary or from both the sources depending upon the objective/s of the investigation

Classification and Presentation : Once data are collected , researcher has to arrange them in a format from which they would be able to draw some conclusions. The arrangement of data in groups according to some similarities is known as classification.

Tabulation is the process of presenting the classified data in the form of table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. Classified and Tabulated data can be presented in diagrams and graphs to facilitate the understanding of various trends as well as the process of comparison of various situations.

Analysis of Data : It is the most important step in any statistical enquiry . Statistical analysis is carried out to process the observed data and transform it in such a manner as to make it suitable for decision making.

Interpretation of Data : After analysing the data, researcher gets information partly or wholly about the population. Explanation of such information is more useful in real life. The quality of interpretation depends more and more on the experience and insight of the researcher.

Self Assessment Questions

1) Complete the following statements

   i) The word statistics has been derived from Latin word .....

   ii) Statistics in plural means .....

   iii) Statistics in singular means .....

   iv) The first step in statistics is .....

   v) The last step in statistics is .....

2) Tick (√) the correct answer

   Statistical data are:

   i) Aggregates of facts

   ii) Unsystematic data

   iii) Single or isolated facts or figure

   iv) None of these

3) Which one of the following statement is true for statistics in singular sense?

   i) Statistics are aggregate of facts.

   ii) Statistics are numerical facts.

   iii) Statistics are collected in a systematic manner.

   iv) Statistics may be called the science of counting.

1.3 TYPES OF STATISTICS

After knowing the concept and definition of statistics, let us know the various types of statistics.

Though various bases have been adopted to classify statistics, following are the two major ways of classifying statistics: (i) on the basis of function and (ii) on the basis of distribution.

1.3.1 On the Basis of Functions

As statistics has some particular procedures to deal with its subject matter or data, three types of statistics have been described.

A) **Descriptive statistics:** The branch which deals with descriptions of obtained data is known as descriptive statistics. On the basis of these descriptions a particular group of population is defined for corresponding characteristics. The descriptive statistics include classification, tabulation measures of central tendency and variability. These measures enable the researchers to know about the tendency of data or the scores, which further enhance the ease in description of the phenomena.

B) **Correlational statistics:** The obtained data are disclosed for their inter correlations in this type of statistics. It includes various types of techniques to compute the correlations among data. Correlational statistics also provide description about sample or population for their further analyses to explore the significance of their differences.

C) **Inferential statistics:** Inferential statistics deals with the drawing of conclusions about large group of individuals (population) on the basis of observations of few participants from them or about the events which are yet to occur on the basis of past events. It provide tools to compute the probabilities of future behaviour of the subjects.

1.3.2 On the Basis of Distribution of Data

Parametric and nonparametric statistics are the two classifications on the basis of distribution of data. Both are also concerned to population or sample. By population we mean the total number of items in a sphere. In general it has infinite number therein but in statistics there is a finite number of a population, like the number of students in a college. According to Kerlinger (1968) "the term population and universe mean all the members of any well-defined class of people, events or objects." In a broad sense, statistical population may have three kinds of properties – (a) containing finite number of items and knowable, (b) having finite number of articles but unknowable, and (c) keeping infinite number of articles.

Sample is known as a part from population which represents that particular population's properties. As much as the sample selection will be unbiased and random, it will be more representing its population. "Sample is a part of a population selected (usually according to some procedure and with some purpose in mind) such that it is considered to be representative of the population as a whole".

Parametric statistics is defined to have an assumption of normal distribution for its population under study. "Parametric statistics refers to those statistical techniques that have been developed on the assumption that the data are of a certain type. In particular the measure should be an interval scale and the scores should be drawn from a normal distribution".

There are certain basic assumptions of parametric statistics. The very first characteristic of parametric statistics is that it moves after confirming its population's property of **normal distribution**. The normal distribution of a population shows its symmetrical spread over the continuum of –3 SD to +3 SD and keeping unimodal shape as its mean, median, and mode coincide. If the samples are from various populations then it is assumed to have same variance ratio among them. The samples are independent in their selection. The chances of occurrence of any event or item out of the total population are equal and any item can be selected in the sample. This reflects the randomized nature of sample which also happens to be a good tool to avoid any experimenter bias.

In view of the above assumptions, parametric statistics seem to be more reliable and authentic as compared to the nonparametric statistics. These statistics are more powerful to establish the statistical significance of effects and differences among variables. It is more appropriate and reliable to use parametric statistics

in case of large samples as it consist of more accuracy of results. The data to be analysed under parametric statistics are usually from interval scale.

However, along with many advantages, some disadvantages have also been noted for the parametric statistics. It is bound to follow the rigid assumption of normal distribution and further it narrows the scope of its usage. In case of small sample, normal distribution cannot be attained and thus parametric statistics cannot be used. Further, computation in parametric statistics is lengthy and complex because of large samples and numerical calculations. T-test, F-test, r-test, are some of the major parametric statistics used for data analysis.

Nonparametric statistics are those statistics which are not based on the assumption of normal distribution of population. Therefore, these are also known as distribution free statistics. They are not bound to be used with interval scale data or normally distributed data. The data with non-continuity are to be tackled with these statistics. In the samples where it is difficult to maintain the assumption of normal distribution, nonparametric statistics are used for analysis. The samples with small number of items are treated with nonparametric statistics because of the absence of normal distribution. It can be used even for nominal data along with the ordinal data. Some of the usual nonparametric statistics include chi-square, Spearman's rank difference method of correlation, Kendall's rank difference method, Mann-Whitney U test, etc.

Self Assessment Questions

1) State true/false for the following statements

   i) Parametric statistics is known as distribution free statistics (T/ F)

   ii) Nonparametric tests assume normality of distribution (T/F)

   iii) T test is an example of parametric test (T/F)

   iv) Nonparametric tests are not bound to be used with interval scale. (T/F)

   v) Parametric tests are bound to be used with either interval or ratio scale. (T/F)

   vi) In case of small sample where normal distribution can not be attained, the use of nonparametric test is more appropriate. (T/F)

2) Define the term sample and population with one example each.

   .....

   .....

   .....

   .....

   .....

   .....

   .....

1.4 SCOPE AND USE OF STATISTICS

Statistical applications have a wide scope. Some of the major ones are given below:

Policy planning: To finalise a policy, it requires some data from previous or expected environment that the policy can be effectively utilised with maximum favourable results. For example, in an organisation the previous sales data are analysed to develop future strategies in the field to obtain maximum benefit in terms of product sale.

Management: Statistics is very useful tool in an organisation to view various aspects of work and well being of the employees as well as keeping an eye on the progress trend of the organisation.

Behavioural and Social Sciences: In social sciences where both types (quantitative and qualitative) of information are used, statistics helps the researchers to alter the information in a comprehensive way to explain and predict the patterns of behaviour/ trend. Where the characteristics of the population being studied are normally distributed, the best and statistically important decision about variables being investigated is possible by using parametric statistics or nonparametric statistics to explain the pattern of activities.

Education: If education is intended to be well dispersed and effective in the interest of the population, the characteristics of students, instructor's contents and infrastructure are very important to understand and again statistics enable these characteristics being analysed in context of needs of the nation. Once the parameters of all components are analysed, areas needing more emphasis become obvious.

Commerce and Accounts: Where money matters are involved, it is essential to take extra care to manage the funds properly enabling efforts in various sectors. The cost and benefit analysis helps to decide putting money and regulating it for maximum benefit at minimum cost.

Industries: Statistics is a basic tool to handle daily matters not only in big organisations but also in small industries. It is required, at each level, to keep data with care and look at them in different perspectives to mitigate the expenditure and enable each employee to have his/ her share in the benefit. Psychologists/ personnel officers dealing with selection and training in industries also use statistical tools to differentiate among employees.

Pure sciences and Mathematics: Statistical tools are also instrumental to have precise measures in pure sciences and to see differences on different occasions in various conditions. Statistics itself is a branch of mathematics which helps them understand differences among properties of various applications in mathematics.

Problem solving: Knowing the useful difference between two or more variables enable the individual to find out the best applicable solution to a problem situation and it is possible because of statistics. During problem solving statistics helps the person analyse his/ her pattern of response and the correct solution thereby minimising the error factor.

Theoretical researches: Theories evolve on the basis of facts obtained from the field. Statistical analyses establish the significance of those facts for a particular paradigm or phenomena. Researchers are engaged in using the statistical measures to decide on the facts and data whether a particular theory can be maintained or challenged. The significance between the facts and factors help them to explore the connectivity among them.

1.5 LIMITATIONS OF STATISTICS

Although Statistics has a very wide application in everyday life as well as in Behavioural Sciences, Physical and Natural Sciences, it has certain limitations also. These limitations are as follow :

Statistics deals with aggregate of facts. It cannot deal with single observation. Thus statistical methods do not give any recognition to an object or a person or an event in isolation. This is a serious limitation of Statistics.

Since Statistics is a science dealing with numerical data, it is more applicable to those phenomenon which can be measured quantitatively. However, the techniques of statistical analysis can be applied to qualitative phenomenon indirectly by expressing them numerically with the help of quantitative standards.

Statistical conclusions are true only on the average . Thus, statistical inferences may not be considered as exact like inferences based on Mathematical laws.

1.6 DISTRUST AND MISUSE OF STATISTICS

Sometimes irresponsible, inexperienced people use statistical tools to fulfill their self motives irrespective of the nature and trend of the data. Because of such various misuses of statistical tools sometimes called an unscrupulous science. There are various misgivings about Statistics . These are as follows :

“Statistics can prove anything”

“Statistics is an unreliable science”

“There are three types of lies , namely, lies, damned lies, and statistics.”

“An ounce of truth will produce tons of Statistics “

Therefore care and precautions should be taken care for the interpretation of statistical data. “ Statistics should not be used as a blind man uses a lamp-post for support instead of illumination”

There are many other fields like, agriculture, space, medicine, geology, technology, etc. where statistics is extensively used to predict the results and find out precision in decision.

Self Assessment Question

1) Write three application of statistics in daily life.

.....

.....

.....

2) List atleast two misuses of statistics.

.....

.....

.....

.....

.....

.....

.....

.....

1.7 LET US SUM UP

In present era people must have some knowledge of statistics. In singular sense, it means statistical methods which include collection, classification, analysis and interpretation of data. In plural sense, it means quantitative information called data. Descriptive, correlational and inferential statistics are three different type of statistics on the basis of their functions. On the other hand, parametric and non parametric are other types of statistics on the basis of the nature of distribution. Statistics has application in almost in all branches of knowledge as well as all sphere of life. Inspite of its wide applicability, it has certain limitations too. Some times inexperienced people misuse statistics to fulfill their own motives.

1.8 UNIT END QUESTIONS

1) What do you mean by statistics? Define its various types with the help of examples of daily life.
2) "Statistical methods are most dangerous tools in the hand of inexpert." Discuss briefly
3) Define following concepts:
   i) Descriptive statistics
   ii) Inferential statistics
   iii) Parametric statistics
   iv) Non parametric statistics
4) Comments on the following statements in two or three lines with reasons:
   i) Statistics in singular sense implies statistical methods.
   ii) Statistics and statistic implies same thing.
   iii) Statistics may rightly be called the science of averages.
   iv) There are lies, damn lies and statistics. Give three examples of misuse of statistics.
5) Write a note on the limitations of statistics.

1.9 GLOSSARY

Statistics in singular sense : In singular sense, it means scientific methods for collection, presentation, analysis and interpretation of data.

Statistics in plural sense : In plural sense it means a set of numerical scores known as statistical data.

Correlational statistics : The statistics which speaks about one or more than one variable's positive or negative magnitude of relationship.

Descriptive statistics : The statistics which describes the tendency or variance of the scores in a distribution.

Inferential statistics : The statistics that enable the researchers to have some conclusions about population or events on the basis of past or observed observations.

Non parametric statistics : The statistics free from the assumptions of normal distribution.

Parametric statistics : The statistics based on assumptions of normal distribution

Statistics : The branch of mathematics that deals with inferring the chances of a particular pattern of population or events on the basis of observed patterns..

1.10 SUGGESTED READINGS

Asthana H.S, and Bhushan, B.(2007) *Statistics for Social Sciences* (with SPSS Applications). Prentice Hall of India

B.L.Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Gupta, S.C.(1990) *Fundamentals of Statistics*. Himalaya Publishing House, Mumbai